Social Network Analysis

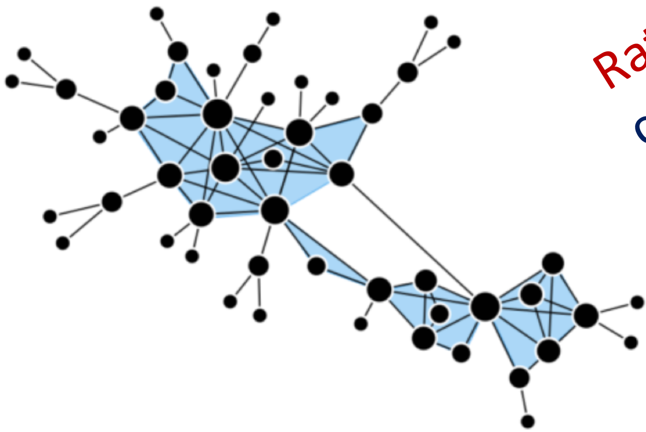
Other Analytics

Clustering coefficient

What is the Clustering coefficient?

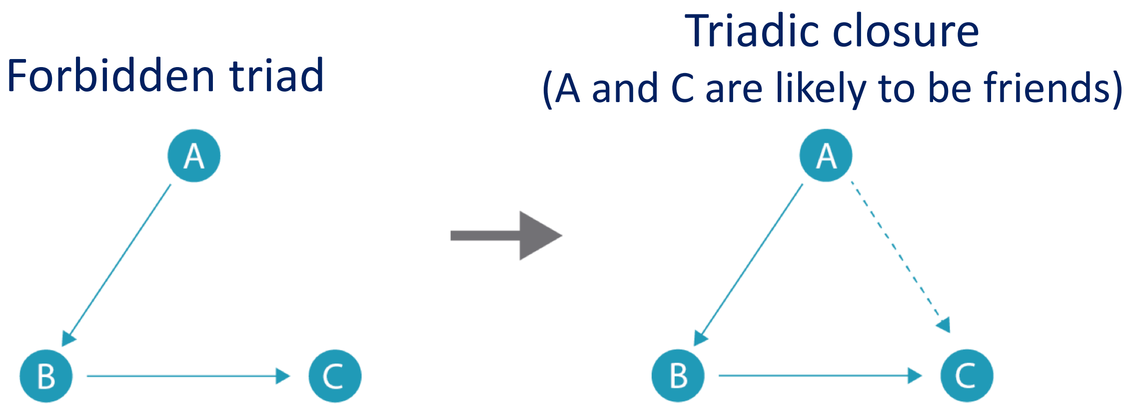
Local clustering coefficient [edit]

The **local clustering coefficient** of a vertex (node) in a graph quantifies how close its neighbours are to being a clique (complete graph). Duncan J. Watts and Steven Strogatz introduced the measure in 1998 to determine whether a graph is a small-world network.



Rationale: how strongly connected is the network locally / general indication of the graph's tendency to be organized into clusters

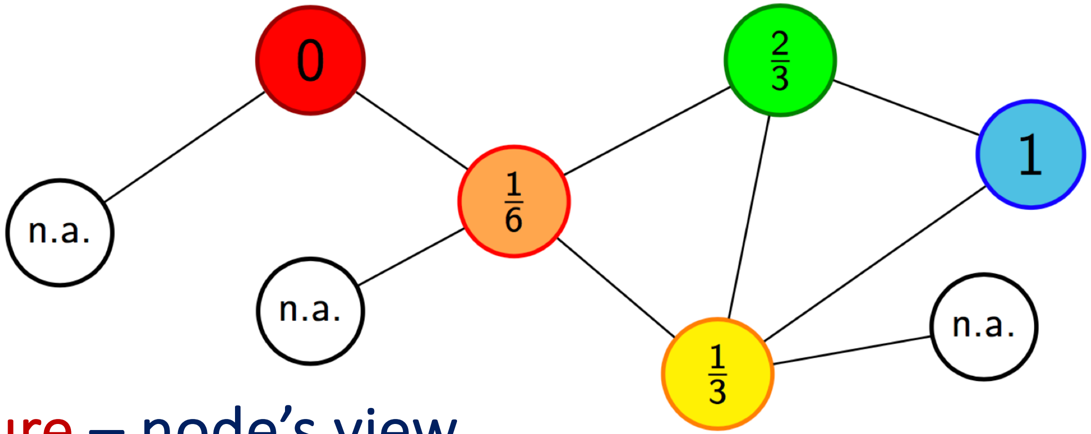
Triadic closure



Triadic closure

- A and C are likely to have the opportunity to meet because they have a common friend B
- The fact that A and C is friends with B gives them the basis of trusting each other
- B may have the incentive to bring A and C together, as it may be hard for B to maintain disjoint relationships

Clustering coefficient and triadic closure



A measure for triadic closure – node's view

- Clustering coefficient C_i
- Counts the fraction of pairs of neighbours which form a triadic closure with node *i*

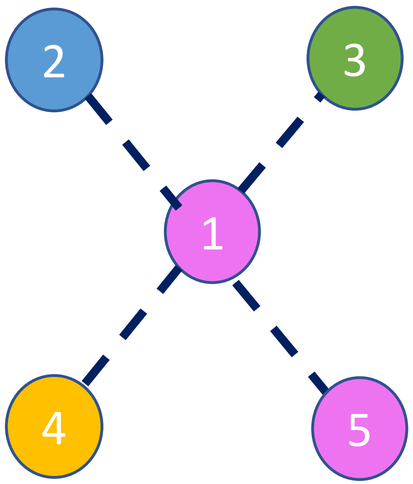
$$C_i = \frac{1}{|\mathcal{N}_i|(|\mathcal{N}_i|-1)} \sum_{\substack{(j,k)\in\mathcal{N}_i^2 \\ j\neq k}} \mathrm{tc}_{i,j,k}$$

where $tc_{ijk} = 1$ if the triplet (i,j,k) forms a triadic closure, and zero otherwise

Examples

not connected
neighbourhood

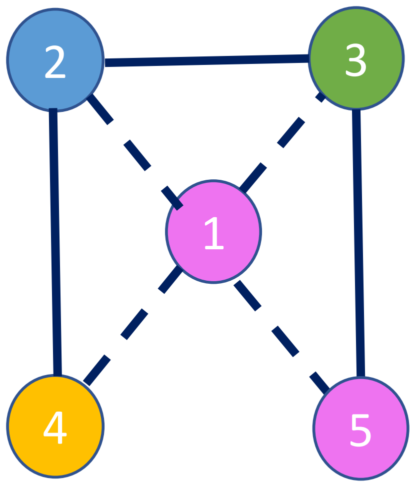
$C_1 = 0$



$\langle C \rangle = 0$

weakly connected
neighbourhood

$C_1 = \frac{1}{2} = 3 / (4x3/2)$

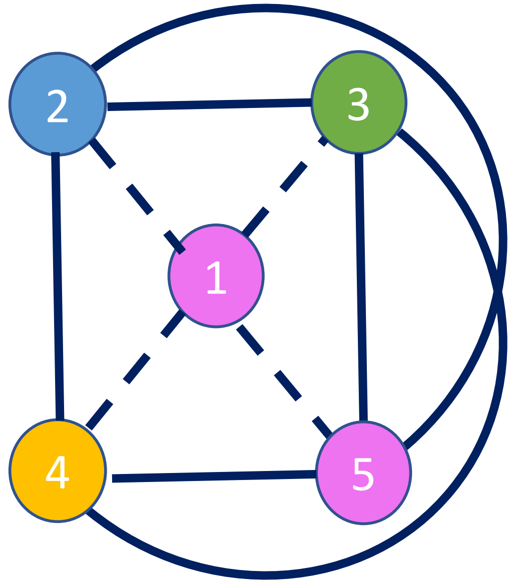


$C_2 = C_3 = \frac{2}{3},\ C_4 = C_5 = 1$

$\langle C \rangle = 0.766$

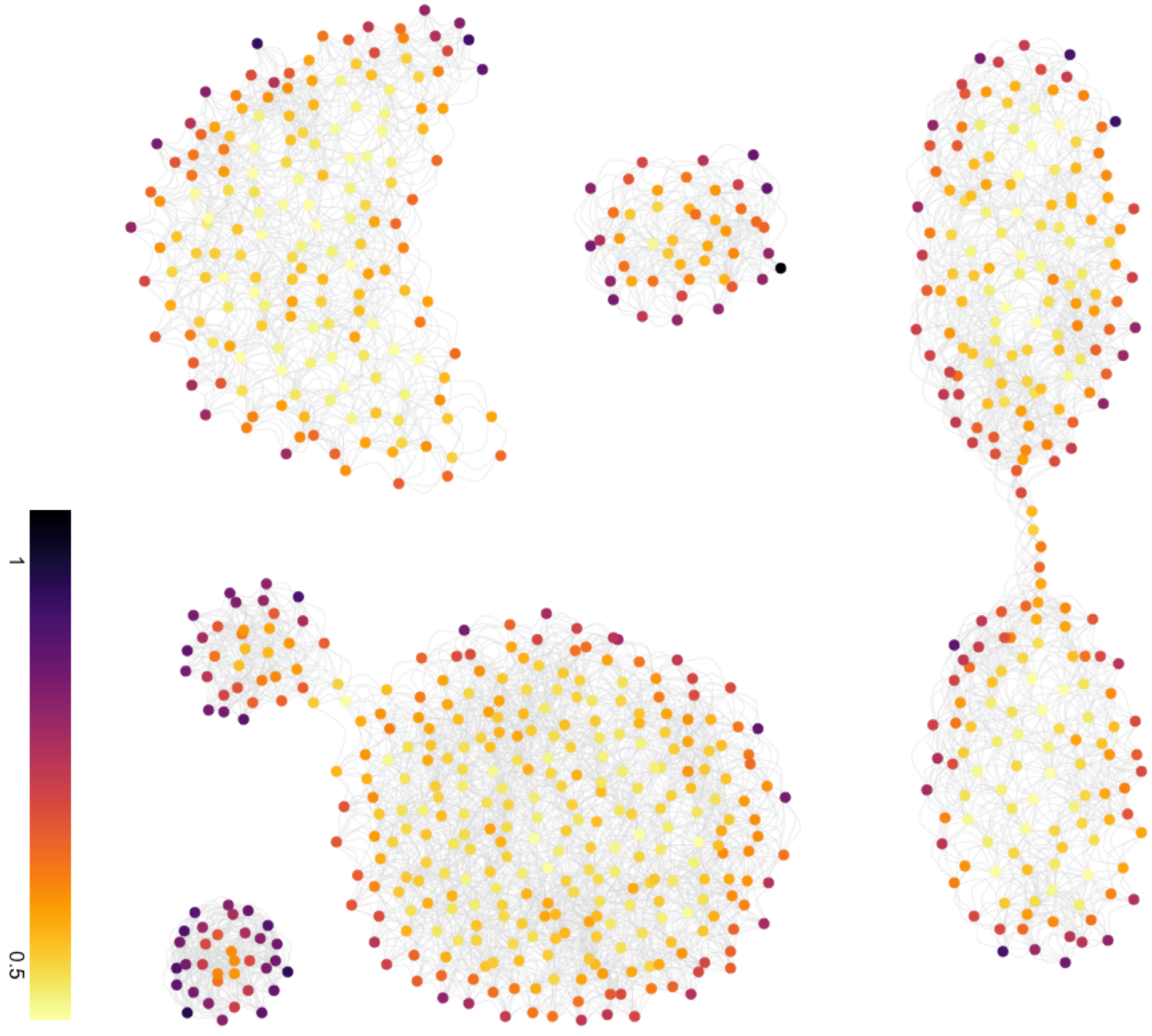
strongly connected
neighbourhood

$C_1 = 1 = 6 / (4x3/2)$



$\langle C \rangle = 1$

Visual example 1



Visual example 2

But clustering coefficient is generally hard to see and visual interpretation is considered unreliable

Robustness

A.L. Barabási, Network science, http://barabasi.com/networksciencebook

Ch.8 “Network robustness”

Network robustness

- We are interested in network robustness to failures
- Want to understand how real networks work under imperfect conditions/malfunctioning

  e.g., why some mutations lead to diseases (biology & medicine)

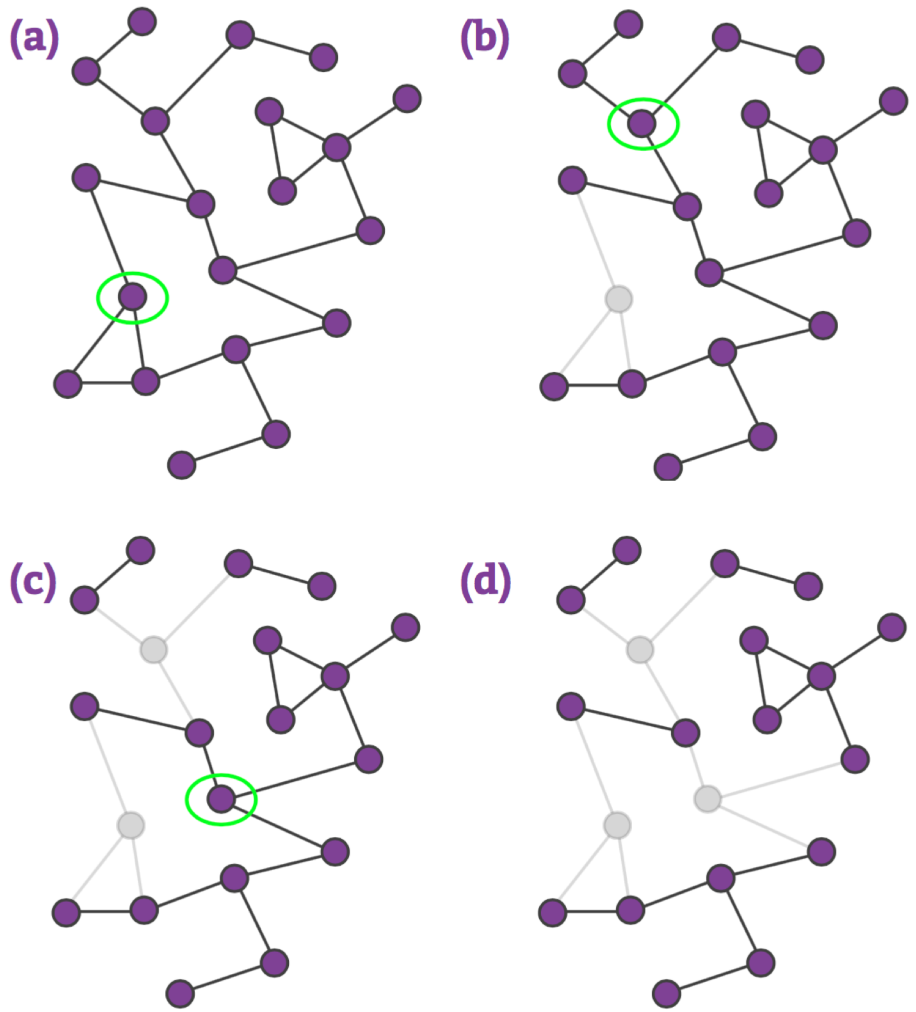
  stability of social networks to disruptive events (war, famine, etc)

  robustness to occasional failures in the www

Oak, Quercus Robur → robust

Network robustness

- Would the network still "work" in the presence of missing nodes?
- Failures can lead to either just isolating nodes or breaking the whole network apart
- What is the limit/phase transition?



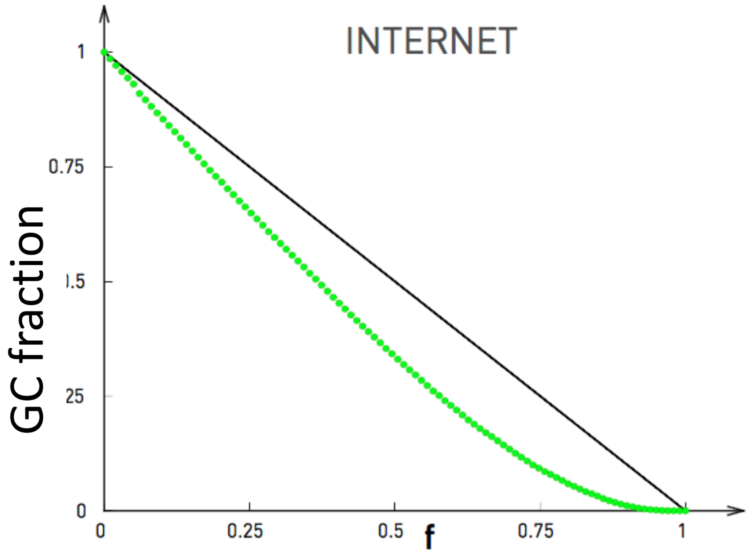
Applications

This can serve to identify:

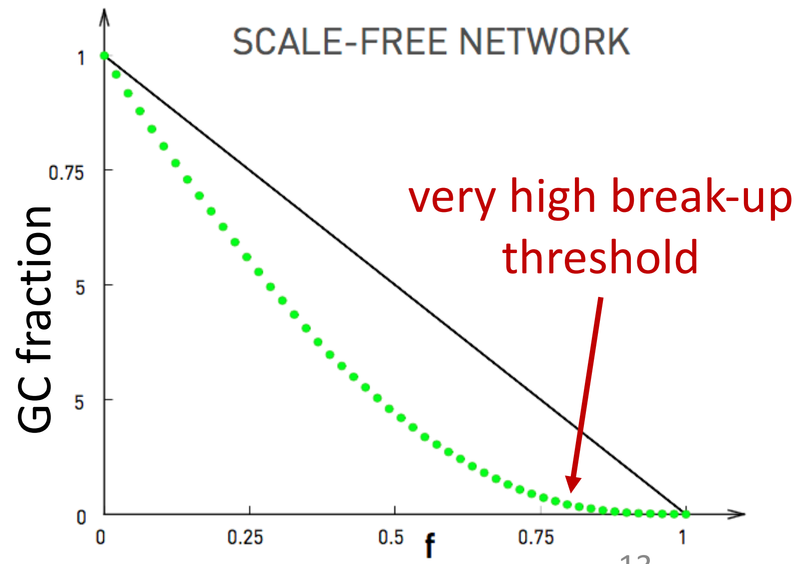
- robustness of air transportation under random strikes
- robustness of social contacts even when someone is off
- possibility of destroying of criminal/terror networks
- eradication of an epidemics

Robustness of scale-free nets

- Robustness of the Internet due to scale-free properties
- Nodes linked to the GC after random removal with rate *f* → still large if *f<1*



- Experiments aligned with a scale-free model
- Reason: random removal of (many) hubs is very unlikely



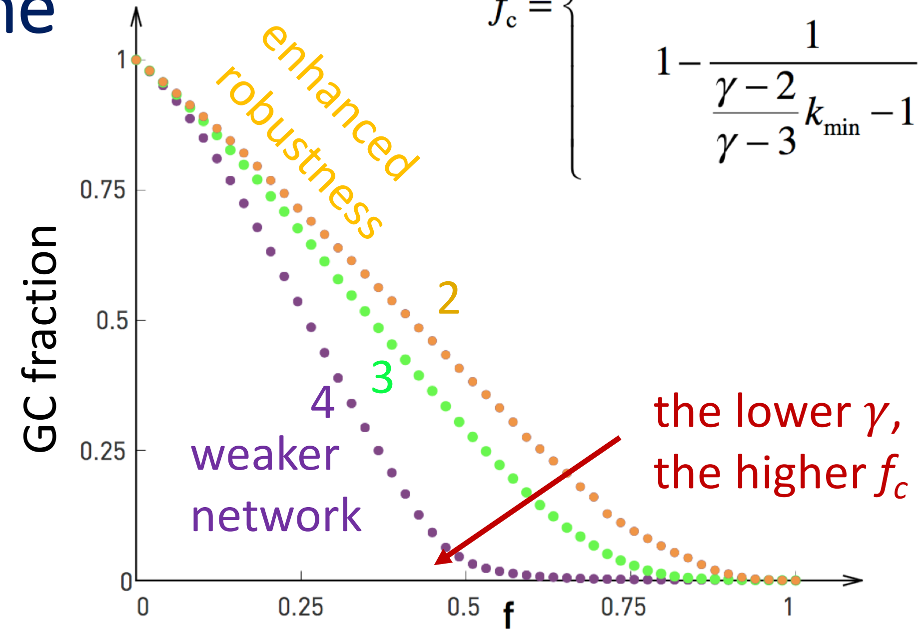
MiME.

Breaking point in scale-free nets

- Assume node removal at rate f
- The inhomogeneity ratio is $\kappa = \langle k^2 \rangle / \langle k \rangle$,
  e.g., in random networks $\kappa = 1 + \langle k \rangle$
- The breaking point is

$f_c = 1 - 1/(\kappa - 1)$ which solely depends on the degree distribution

$$f_c = \begin{cases} 1 - \dfrac{1}{\dfrac{\gamma-2}{3-\gamma} k_{\min}^{\gamma-2} k_{\max}^{3-\gamma} - 1} & 2 < \gamma < 3 \\[2ex] 1 - \dfrac{1}{\dfrac{\gamma-2}{\gamma-3} k_{\min} - 1} & \gamma > 3 \end{cases}$$



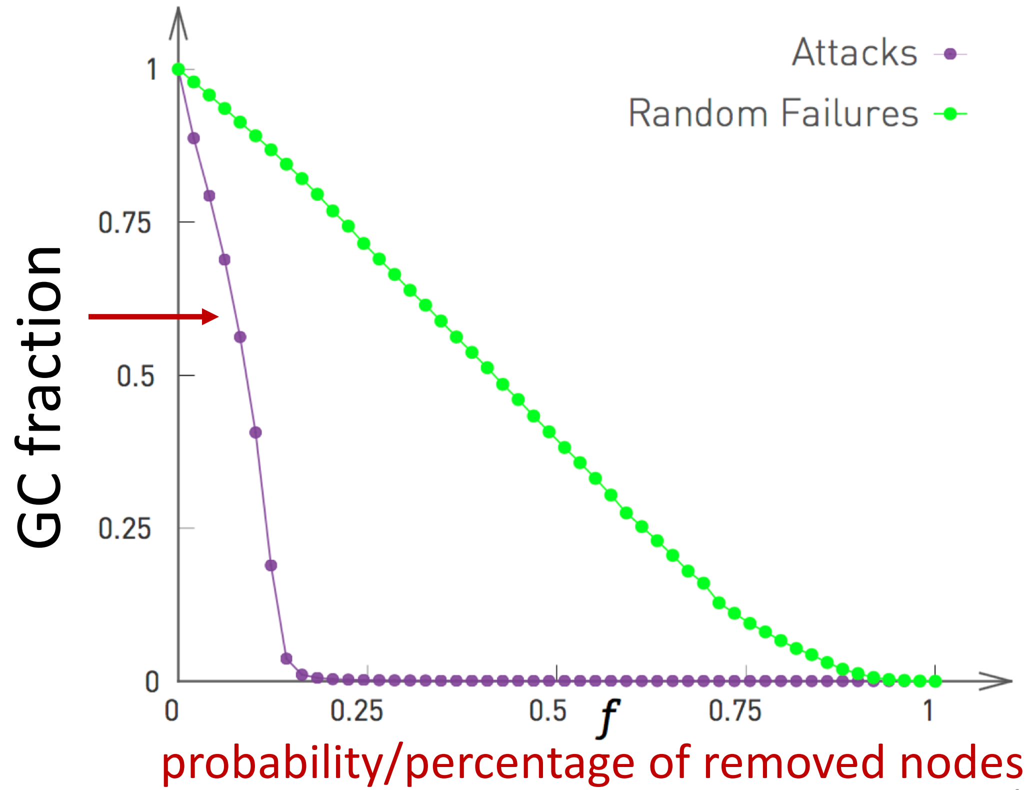
Some implications

- networks with big hubs (causing wide deviations from $\langle k \rangle$) are hard to die
- in random networks $f_c = 1 - 1/\langle k \rangle$, i.e., large average degrees strengthen the network
- in scale-free networks the exponent γ sets the network robustness

Attack tolerance

- What if removals are not by chance, but caused by an adversary with sufficient insights on our network?

an adversary would remove all hubs first, i.e., it removes nodes in decreasing order of their degree



Fragility of scale-free nets

- Scale-free networks are not very robust to targeted attacks exactly because they have vulnerable hubs
- Recall that $f_c = 1 - 1/(\kappa - 1)$ meaning that robustness depends on κ, and removing hubs reduces κ
  - good news in medicine (vulnerability of bacteria) ☺
  - bad news for the Internet ☹

MIME.

Breaking point in scale-free nets

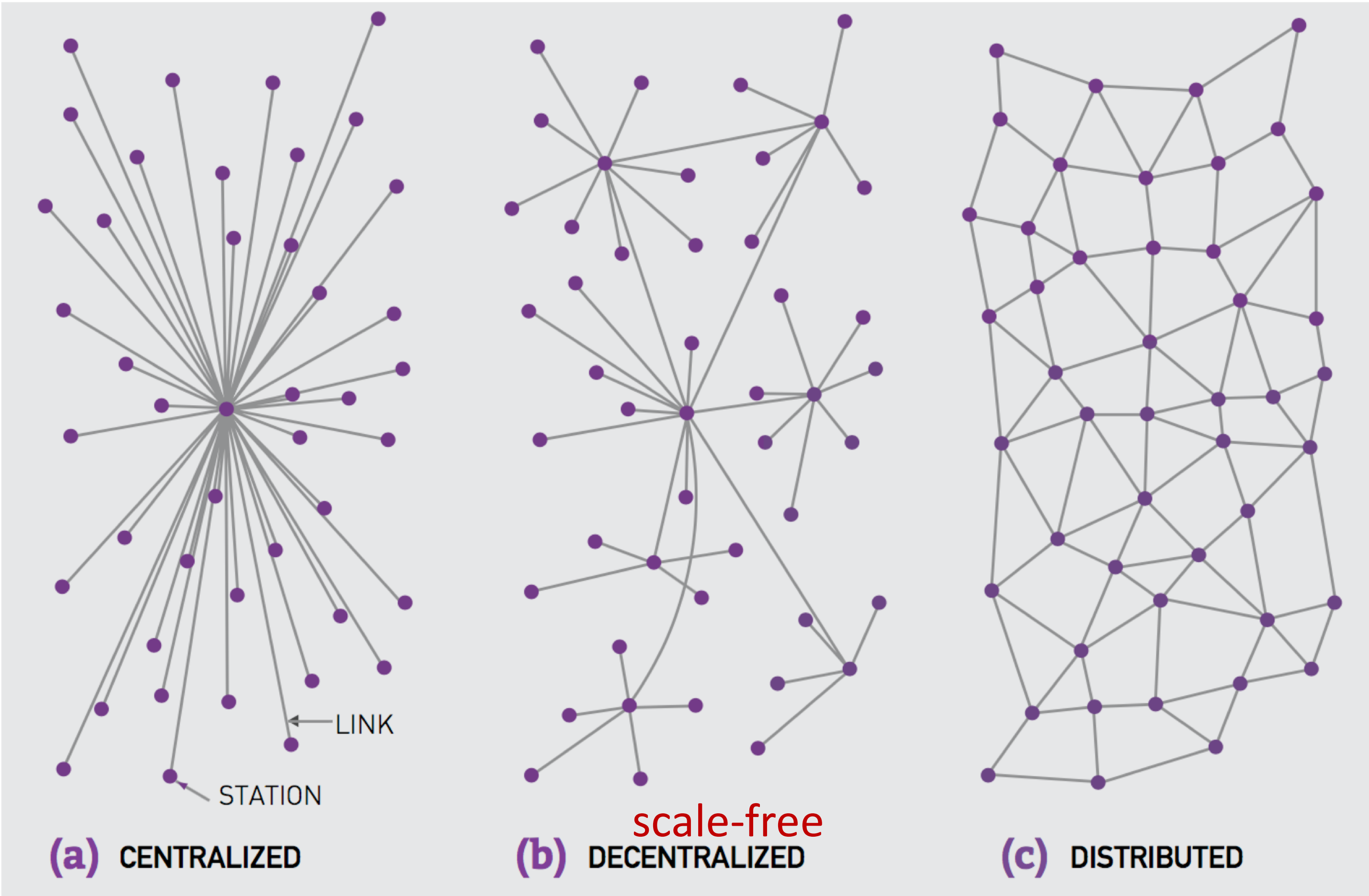
estimated value

NETWORK	RANDOM FAILURES (REAL NETWORK)	RANDOM FAILURES (RANDOMIZED NETWORK)	ATTACK (REAL NETWORK)
Internet	0.92	0.84	0.16
WWW	0.88	0.85	0.12
Power Grid	0.61	0.63	0.20
Mobile-Phone Call	0.78	0.68	0.20
Email	0.92	0.69	0.04
Science Collaboration	0.92	0.88	0.27
Actor Network	0.98	0.99	0.55
Citation Network	0.96	0.95	0.76
E. Coli Metabolism	0.96	0.90	0.49
Yeast Protein Interactions	0.88	0.66	0.06

Not robust to random failures (exponential degree distribution)

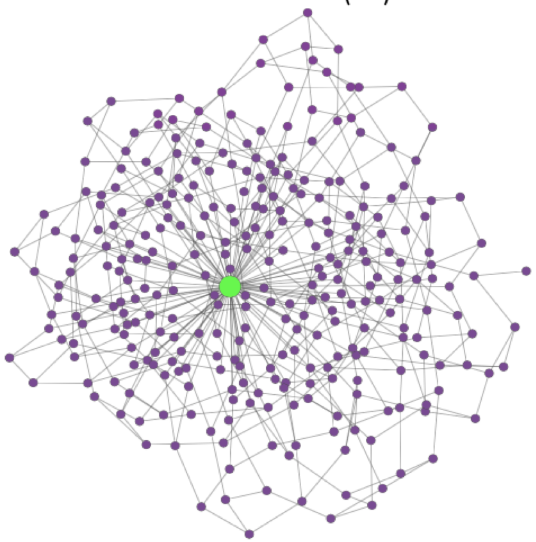
Optimizing robustness

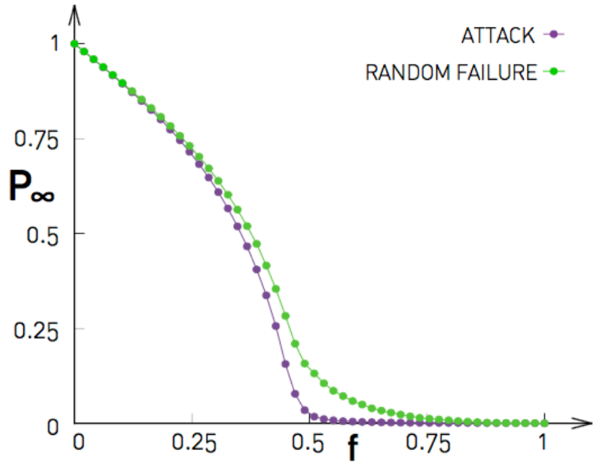
An early attempt by Paul Baran [1959]



Optimizing robustness

$\langle k \rangle = 3$

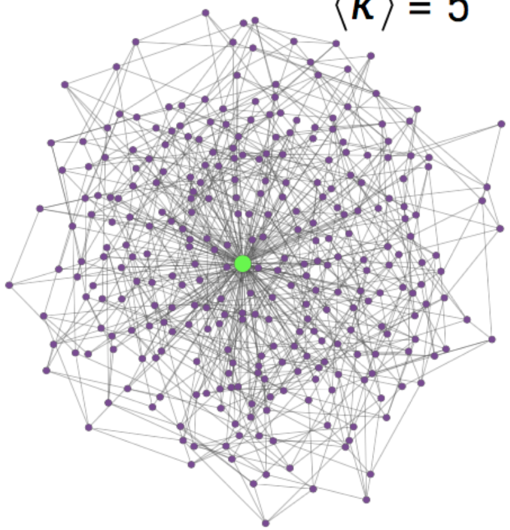




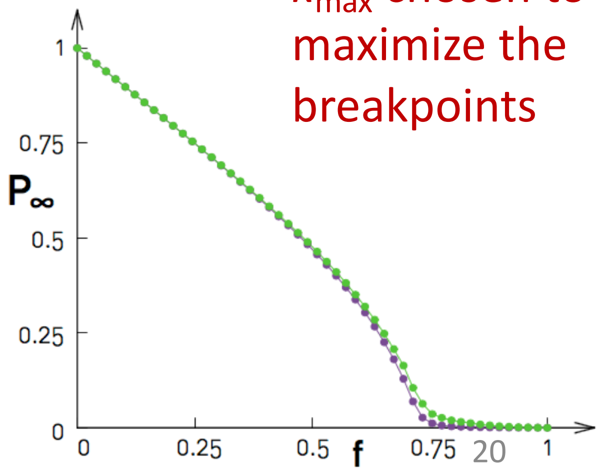
The best option is a **bimodal** distribution

$$p_k = r\,\delta_{k_{\mathrm{max}}} + (1-r)\,\delta_{k_{\mathrm{min}}}$$

$\langle k \rangle = 5$



r = 1/N
k_{max} chosen to maximize the breakpoints

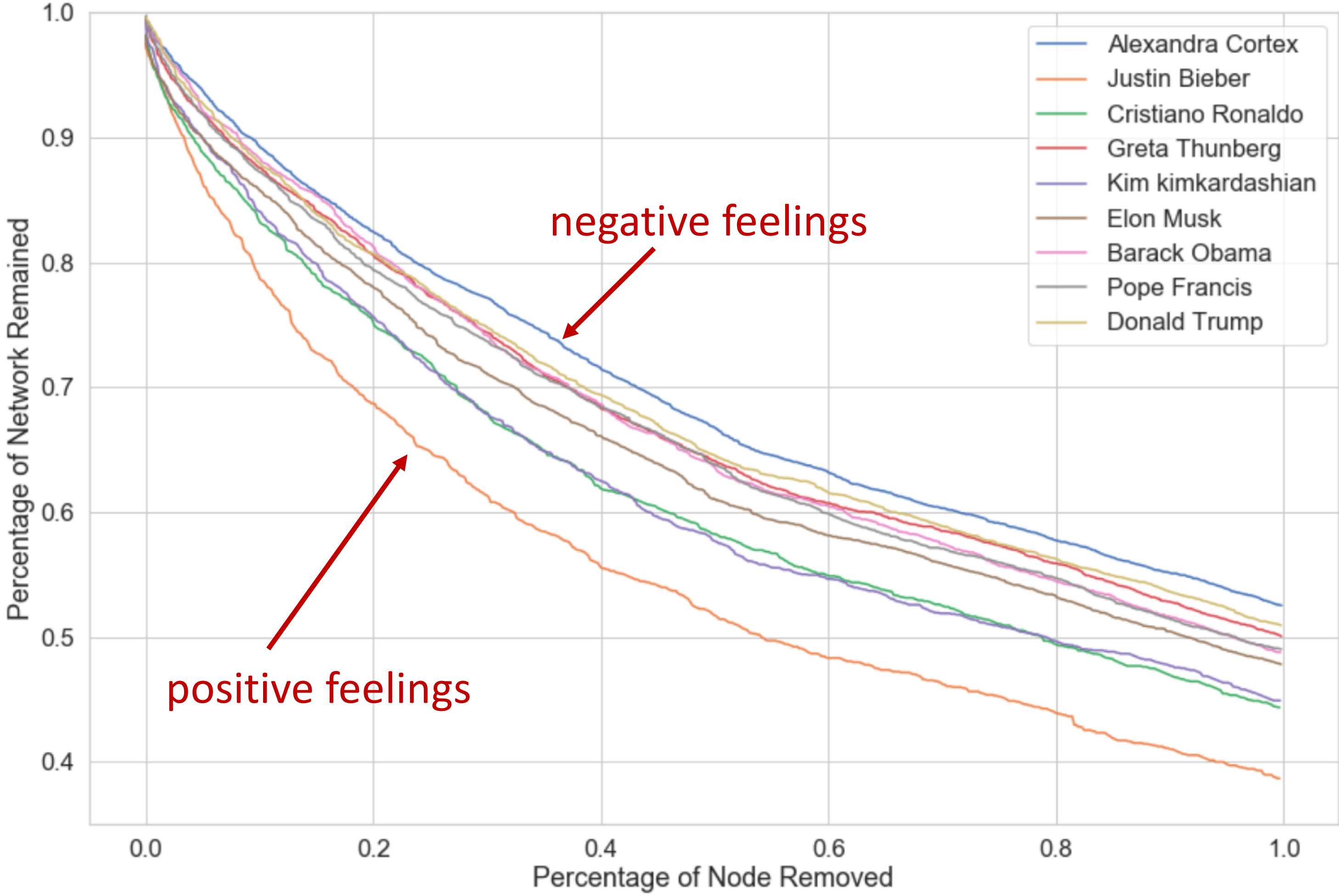


MIME.

Network analysis of Tweets' sentiment

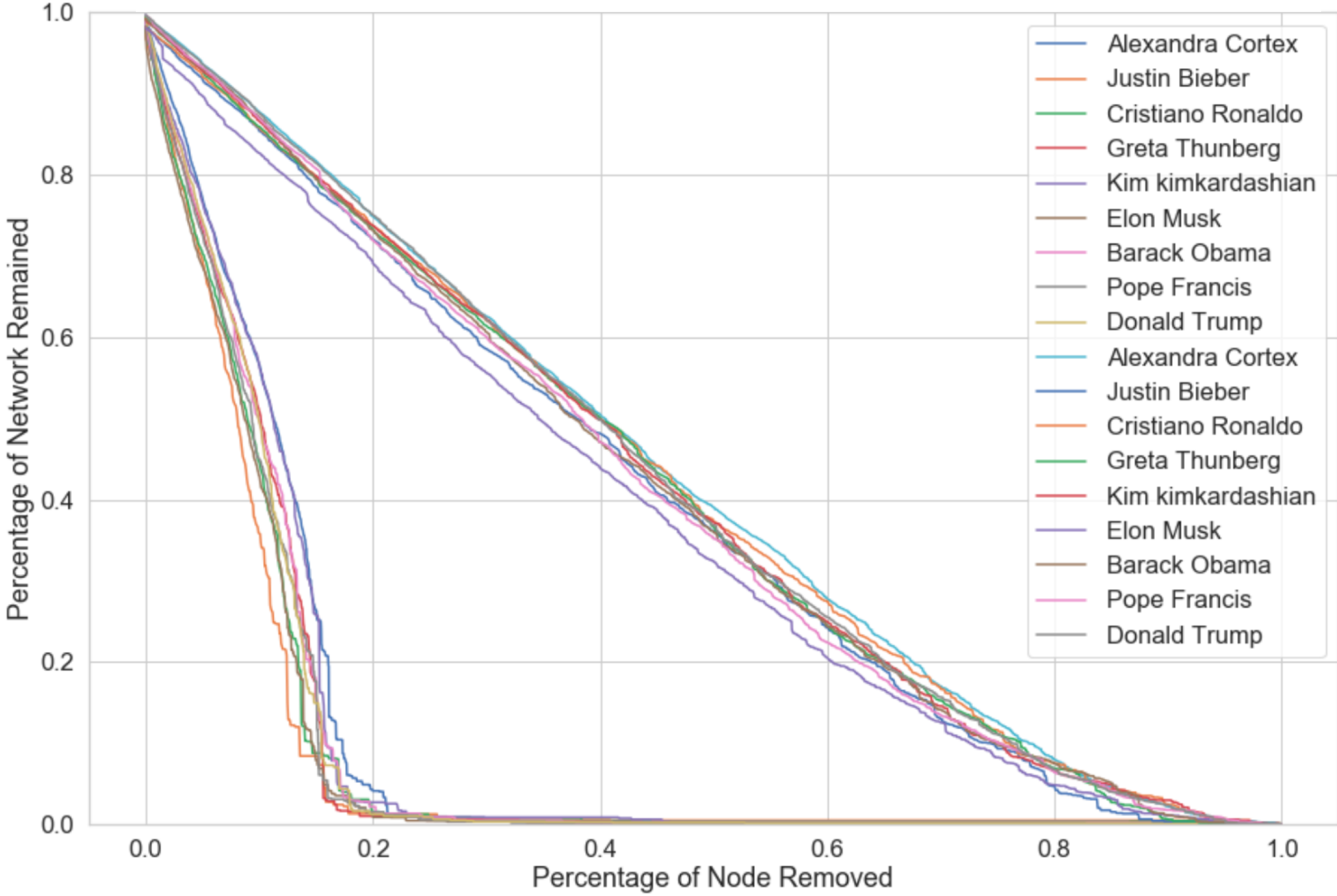
Salvatore Romano, Alberto Zancanaro, Enrico Lanza, Carlo Facchin

Robustness of original network to positive node removal



Network analysis of Tweets’ sentiment

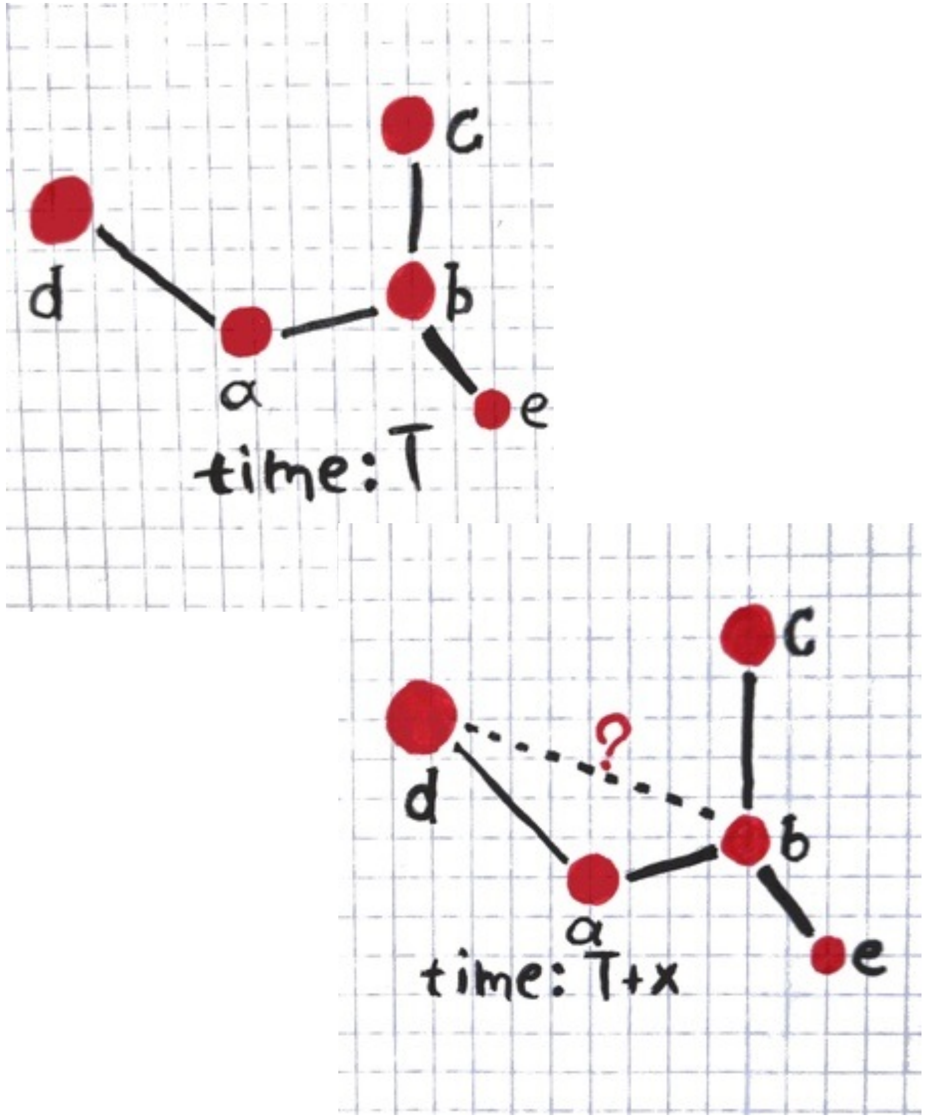
Robustness of original network to positive node removal



Link Prediction

The link prediction task

Given a graph at time T, can we output a **ranked list** of links that are **predicted** to appear in the graph at time T+x ?



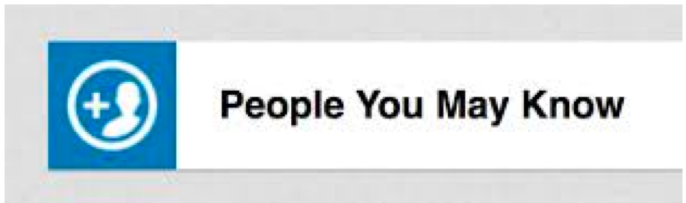
idea

We can build the list by using a measure of **similarity**/proximity between nodes

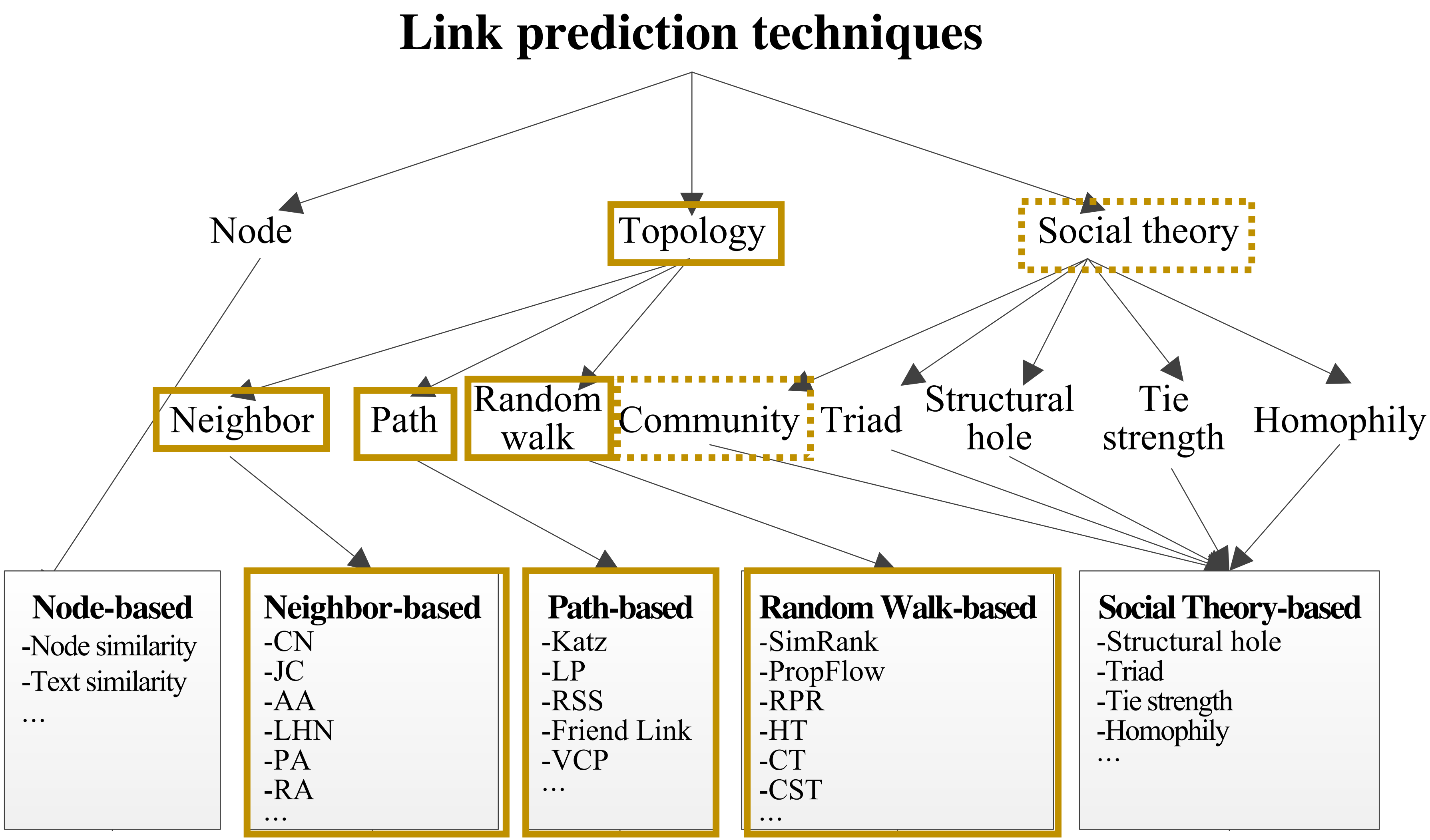
The link prediction task

Applications:

- Recommendation in **social** networks
- Finding experts and collaborations in **academic** social networks
- Reciprocal **relationships** prediction
- Network **completion** problem
- Social **tie** prediction
- …



The link prediction task



MIME

Neighbour based techniques

These **local** techniques are modification of a simple idea

Common neighbours - CN

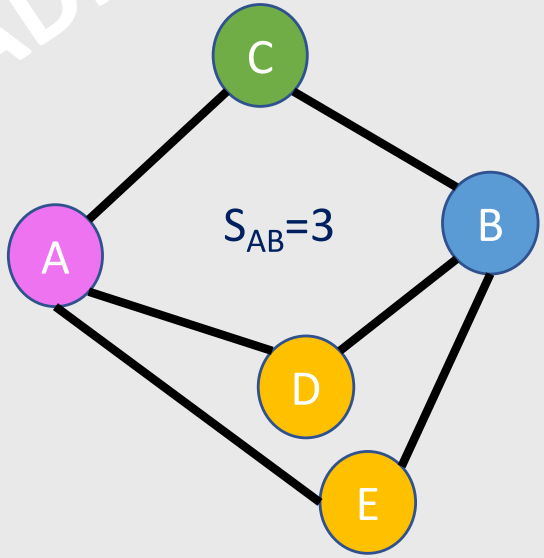
The more neighbours in common, the more likely the link to appear

intersection

$$S_{CN}(i,j) = |N_i \cap N_j|$$

(the set of) neighbours of *j*

SIMPLE TRIADIC CLOSURE



$$\mathbf{S_{CN} = A \cdot A}$$

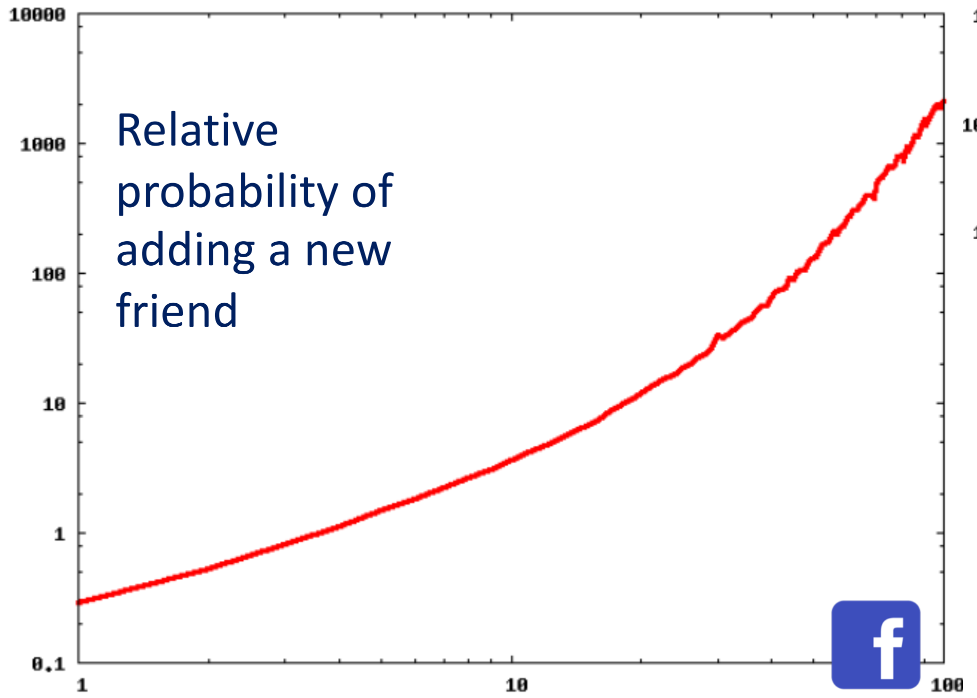
binary adjacency matrix of an undirected network

MIME.

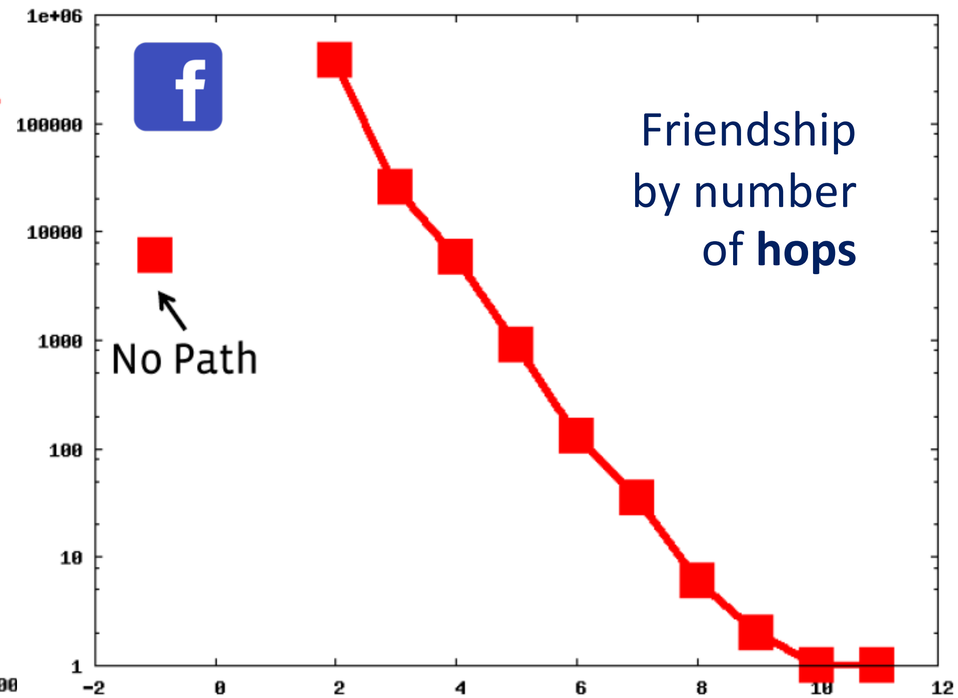
Neighbour based techniques

more **mutual friendships** help in becoming a friend

95% of the new friendships in facebook are **friend-of-a-friend**

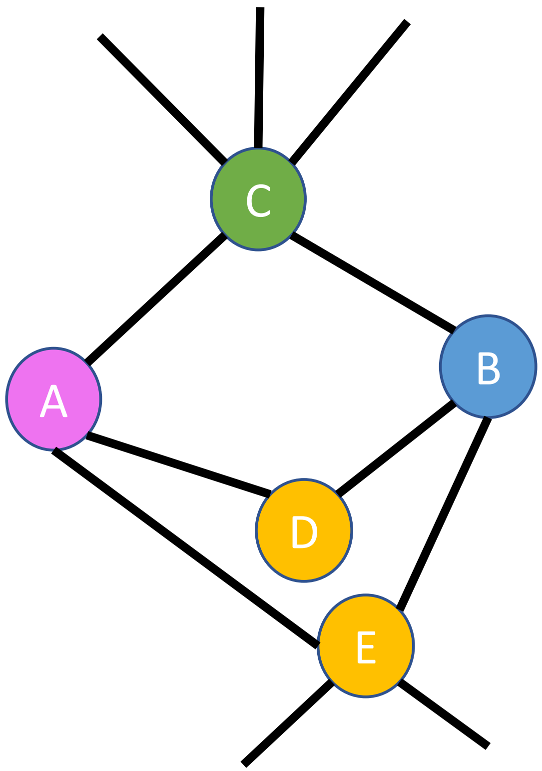


# of friends in common



# of hops

Neighbour based techniques



S_{AB} = 1/5 + 1/2 + 1/4 = 19/20

Resource allocation - RA

Punishes more heavily the **high-degree** common neighbours

$$S_{RA}(i,j) = \sum_{k \in N_i \cap N_j} 1 / |N_k|$$

… but very many variations exist

Path based techniques

These **global** techniques are a generalization of CN to take into account the (very many) paths of **length** $\ell \geq 2$

Kats

of paths of length ℓ between nodes *i* and *j*

$$\mathbf{S}_{\text{Katz}} = \sum_{\ell \geq 1} \beta^{\ell} \, \mathbf{A}^{\ell}$$

damping factor (weights more shorter paths), it needs to be sufficiently **small** $0 < \beta < 1$

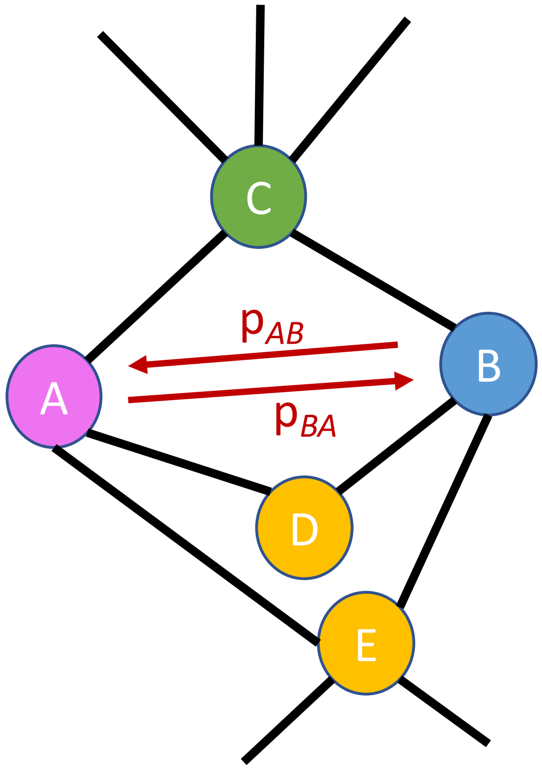
Local path - LP

$$\mathbf{S}_{\text{LP}} = \mathbf{A}^2 + \beta \, \mathbf{A}^3$$

MiME.

Random walk based techniques

These **global** techniques exploit the Local PageRank value



random walk with restart

teleportation to node *i*

$$\mathbf{p}_i = c\,\mathbf{M}\,\mathbf{p}_i + (1-c)\,\mathbf{e}_i$$

Random walk with restart - RWR

$$S_{RWR}(i,j) = p_{ij} + p_{ji}$$

MiME.

Ingredients Networks - Pasta

Elena Camuffo, Laura Crosara, Matteo Moro

pairings		CN	AA	RA	KA	LP	RW
Nutmeg	Fresh chilli	x			x	x	
Liquid fresh cream	Carrots	x			x	x	
Tomato sauce	Pine nuts	x			x	x	
Butter	Mussels	x			x	x	
Salt	Nduja						x
Pig cheek	Pumpkin		x				
Pig cheek	Ricotta cheese	x					
Sausage	Pecorino			x			
Whole milk	Beans			x			
Whole milk	Onions golden		x		x	x	

ITALY

pairings		CN	AA	RA	KA	LP	RW
cheese	sesame	x			x	x	
macrophyll	bean			x			
salt	sweet sauce		x				x
cabbage	lemon			x			
lemon	mushrooms maitake			x			
chicken	vegetables			x			
cabbage	cheese parmigiano			x			
consomme	perilla	x			x	x	
egg	lemon	x		x	x	x	
bacon	vinegar	x			x	x	

JAPAN

pairings		CN	AA	RA	KA	LP	RW
fresh cream	chili	x		x	x	x	
black pepper	potato	x					
spices	bacon	x			x	x	
carrots	nuts		x				
canned tomatoes	pesto	x			x	x	
carrots	pesto		x				
salt	pig cheek						x
lemon juice	chicken broth		x				
rosemary	chicken broth			x			
fresh cream	sugar	x		x	x	x	

TAIWAN

Ingredients Networks - Pasta

ITALY

New Ingredient	Recipe
Black pepper	Durum wheat semolina, Water, Ricotta salata, Eggplant, Garlic, Vine-ripened tomatoes, Basil, Salt, Extra virgin olive oil
Vegetable broth	Semolina durum whole wheat, Water, Fresh onion, Mushrooms, Bacon, Cannellini beans, Rosemary, Extra virgin olive oil, Black pepper, Salt
apple	onion, anchovies, water, olive oil
Brandy	Chicken breast, Noodles, Potatoes, Snow peas, Carrots, Celery, Mushrooms, Leeks, Water, Fresh ginger, Parsley, Extra virgin olive oil, Black pepper, Salt
Almonds	streaky pork, durum wheat semolina, water, minced garlic, plum, cauliflower, mushroom, soft-boiled eggs, rice wine, salt, flour

TAIWAN

New Ingredient	Recipe
mushroom	onion, meat, red wine, concentrated tomato paste, chicken broth, bay leaves, sugar, salt, durum wheat semolina, water, cheese, fresh thyme, black pepper
chia	streaky pork, durum wheat semolina, water, minced garlic, plum, cauliflower, mushroom, soft-boiled eggs, rice wine, salt, flour
cheese	durum wheat semolina, water, bacon, asparagus, shrimp, garlic, black pepper, rose salt, paprika, parsley leaf, cheese
basil leaves	durum wheat semolina, water, onion, cream, chicken breast, squid
avocado	durum wheat semolina, water, bacon, large tomatoes, green pepper, mushroom, cheese, ketchup, salt, black pepper

JAPAN

New Ingredient	Recipe
consomme	durum wheat semolina, water, salmon, olives oil
tomato	onion, bacon, garlic, olives oil, cream, salt, cheese, durum wheat semolina, water, juice, nut
soy sauce	chicken, salt, durum wheat semolina, water, avocado, clams, mayonnaise, onion, cod roe
onion	durum wheat semolina, water, saury, salt
pepper	durum wheat semolina, water, salmon, olives oil

MIME.

Questions ?